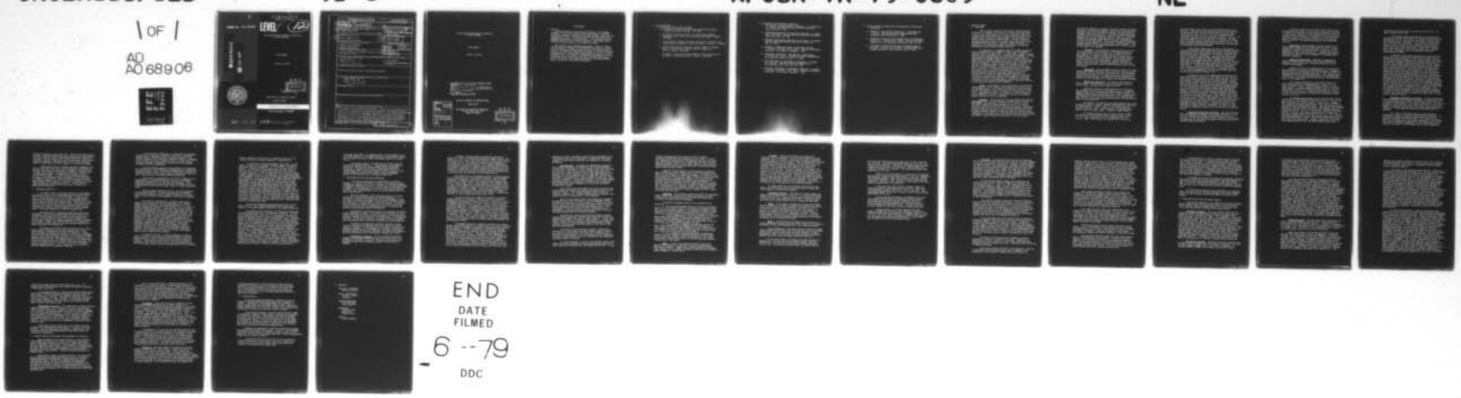
AD-A068 906 ARIZONA STATE UNIV TEMPE DEPT OF EDUCATIONAL TECHNOLOGY F/G 5/9
RULE LEARNING AND SYSTEMATIC INSTRUCTION IN PILOT TRAINING.(U)
MAR 79 V S GERLACH AFOSR-76-2900

UNCLASSIFIED TD-3 AFOSR-TR-79-0609 NL

1 OF 1
AD
A068906



END
DATE
FILMED
6 -- 79
DDC

CENTER FOR INSTRUCTIONAL AND
TRAINING TECHNOLOGY

AFOSR-TR- 79-0609

LEVEL II

12

Rule Learning and Systematic Instruction in Pilot Training

\- Final Report -

Vernon S. Gerlach

USAF Office of Scientific Research

Grant #76-2900

ARIZONA STATE UNIVERSITY
EDUCATIONAL TECHNOLOGY

Technical ~~Document~~ #3-~~76-2900~~

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 18 19 AFOSR-TR- 79-0609	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) 6 RULE LEARNING AND SYSTEMATIC INSTRUCTION IN PILOT TRAINING.		5. TYPE OF REPORT & PERIOD COVERED 9 Final Scientific Report, 01 Oct 77 - 31 Dec 78,
		6. PERFORMING ORG. REPORT NUMBER TR #3-76-2900
7. AUTHOR(s) 10 Vernon S. Gerlach		8. CONTRACT OR GRANT NUMBER(s) 15 AFOSR-76-2900
9. PERFORMING ORGANIZATION NAME AND ADDRESS Arizona State University College of Education Tempe, Aaizona 85281		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 16 61102F 2313/A2 17 A2
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research (NL) Bolling AFB DC 20332		12. REPORT DATE 11 March 1979
		13. NUMBER OF PAGES 27
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) 12 29 p.		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

14 TD-3

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Four related lines of endeavor were pursued. Central to these activities was the research on algorithmized instruction as a form of rule learning and its effect on the acquisition of complex human behaviors. This theme was implemented in studies on the role of performance objectives in instructional systems design, on the role of self-evaluation and its relationship to performance measurement, on the use of computer models in defining algorithms involving rule-using behavior, and an observation regarding the use of the Pearson product-moment correlation in performance measurement.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE 409 075 UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Rule Learning and Systematic Instruction
in Pilot Training

- Final Report -

Vernon S. Gerlach

Technical Document #3 (AFOSR-76-2900)

March 1979

Department of Educational Technology
Arizona State University
Tempe, AZ 85281

Final Report

A. Overview

The report covers our activities from October 1, 1977 through December 31, 1978, when funding of the project ceased. (Some data collection, analysis, and report writing was continued during the first three months of 1979). The project was entitled "Rule Learning and Systematic Instruction in Pilot Training". It was funded as Project #AFOSR-76-2900, Office of Scientific Research. The Arizona State University account number for the project was 915948.

Four related lines of endeavor were pursued. Central to these activities was the research on algorithmized instruction as a form of rule learning and its effect on the acquisition of complex human behaviors. This theme was implemented in studies on the role of performance objectives in instructional systems design, on the role of self-evaluation and its relationship to performance measurement, on the use of computer models in defining algorithms involving rule-using behavior, and an observation regarding the use of the Pearson product-moment correlation in performance measurement.

B. Work Accomplished

1. The following reports were completed:

    a. Principles for Developing Algorithmic Instruction by Richard F. Schmid and Vernon S. Gerlach. (TR 81201)

    b. The Effects of Self-Evaluation as an Instructional Feedback Strategy by Lou M. Carey, Larry Israelite, and Richard F. Schmid. (TR 81202)

    c. The Relationship between Algorithmic Processes for Instruction and Computer Models by Richard F. Schmid and Vernon S. Gerlach. (TR 81203)

    d. Performance Objectives by Vernon S. Gerlach, Robert C. Haygood, Gary L. Filan, Richard F. Schmid, Dianne L. Wigand, and William V. Hagin. (TR 81230)

    e. An Anomaly in the Pearson Product-Moment Correlation Coefficient by Robert C. Haygood and Vernon S. Gerlach. (TR 81229)

2. Presentation at Professional Conventions:

   a. The efficiency of algorithmized instruction. V.S. Gerlach and R.F. Schmid. American Educational Research Association, Toronto, March 27, 1978.

   b. Determining the worth of instructional objectives. N. Higgins and S. Herrington. American Educational Research Association, Kansas City, April 18, 1978.

   c. Preparing instructional developers for the 1980's. V. Gerlach. American Educational Research Association, Kansas City, April 18, 1978.

   d. Seminar on research and theory in algorithmic learning. V. Gerlach, J. Whitaker, and M. Pearce. American Psychological Association, San Francisco, September 1, 1978.

   e. Algorithms in education: some empirical considerations. V. Gerlach and R. Schmid. American Educational Communications and Technology, Kansas City, April 19, 1978.

   f. The effectiveness of algorithms in instruction. R. Schmidt and V. Gerlach. American Educational Communications and Technology, Kansas City, April 20, 1978.

   g. Structural comparisons in algorithmic instruction. R. Schmid, V. Gerlach, J. Whitaker, and M. Pearce. American Psychological Association, San Francisco, September 1, 1978.

3. Papers Accepted and Scheduled for Presentation at Professional Conventions:

   a. Feedback in algorithmized instruction. S. Herrington and V. Gerlach. American Educational Communications and Technology, New Orleans, March 7, 1979.

   b. An empirical analysis of the characteristics of performance objectives. V. Gerlach and R. Schmid. American Educational Communications and Technology, New Orleans, March 8, 1979.

   c. Algorithmic instructional processes and computer models. R. Schmid, V. Gerlach, and M. Valach. American Educational Research Association, San Francisco, April 12, 1979.

4. Research Summary

a. Algorithms

One of the deficiencies in the research literature on rule learning, as it appears in the form of an instructional algorithm, is the lack of an operational definition of what constitutes effectiveness; under these conditions, it seems hazardous to specify the variables which contribute to effectiveness. We, therefore, conducted a study to assess the effectiveness, in terms of immediate and delayed posttest scores, of an instructional algorithm in three representational forms (prose; flowchart; gradual withdrawal of flowchart) in two availability modes (with recourse to algorithm during practice and posttest; without recourse).

An algorithm is "a procedure which will produce the correct result when applied to any problem of a given class of problems" (Gerlach, Reiser, & Brecke; 1976). In recent years a number of works advocating the algorithmization of instruction have appeared. Although some research has demonstrated that instructional algorithms can produce desired outcomes (Landa, 1973) it still remains unclear as to exactly when an algorithm should be used and in what form algorithms should be presented. Landa (1976) has stated that, at present, the only way to determine whether a given unit of subject matter is amenable to algorithmization is to try to construct the algorithm and, if successful, to test the algorithm with appropriate learners. The present study is an effort to go beyond this kind of primitive empiricism by providing data which can be used in beginning to formulate general principles concerning algorithmization of instruction. The effects of three different forms of an algorithm on learner achievement were observed, both immediately following instruction and after a one-week delay. In addition, we observed the degree to which subjects became dependent on the actual presence of the algorithm during and after instruction.

Design. Three factors, Tax Law Availability, Instructional Representation, and Test Interval were combined factorially to form six treatment groups. Test Interval was varied as a within-subject factor. The design was thus a 2 Availability (with tax law vs without) X 3 Representation (prose vs. flowchart vs. faded flowchart) X 2 Test (immediate vs. delay). Analyses of variance, with repeated measures on Test Interval, were employed.

Materials. The instructional task was adapted from Horabin's Algorithms (1974). The task consisted on generating the solution to tax problems involving the purchase and sale of shares of stock. The authors created two sample problems for each of six possible solutions, from which two sets of problems were created. Instruction on how to apply the tax law took one of three representational forms: the prose condition received a verbal description of the law, which followed an if/then format; the flowchart condition received the same information in a flowchart form, such that each decision (discriminator) was binary, leading to another discriminator, until the terminal solution box (operator) was reached; the faded flowchart was exactly the same as the flowchart treatment, except that for each problem completed, one of the discriminators necessary for the solution of subsequent problems was deleted. The deleted information was made available only if the subject was unable to recall it. Corrective feedback was supplied in separate booklets. The questionnaire following the treatment asked about the learner's strategies and reactions to the instruction. The posttest contained six additional randomly ordered problems of the same type, without feedback. The delayed posttest and the immediate posttest were identical, except for the order of the problems.

Procedure. The experimental sequence consisted of (a) an orientation, (b) a practice session, (c) an immediate posttest, (d) a questionnaire, (e) a one-week delay posttest, and (f) a questionnaire. In addition to completing six practice and six posttest problems, subjects recorded the time in spaces provided at the start and finish of each problem. Following the introduction read aloud by the experimenter, subjects completed all phases at their own pace. During both posttests, only half the subjects from each condition were supplied the tax law for solving the problems.

Data Source. Data consisted of the test scores of 77 undergraduate volunteers from Arizona State University. Eleven subjects were dropped from the experiment for failure to follow procedures, leaving 11 per factorial cell. Subjects were run in groups during normal classroom sessions. Materials were prepared beforehand and shuffled, so that assignment to treatments was completely randomized.

Results and Conclusions. All protocols were scored for number correct, with one point for the dollar amount and one for the tax status. Omissions were counted as errors.

Analyses of variance (3 Representation X 2 Availability) were conducted on the practice and posttest sections. A repeated measures anova using the delayed test data was then performed. The repeated measures anova was conducted twice: first, using groups as represented by the with/without distinction on the immediate posttest, and next, using groups separated according to the Availability factor on the delayed test.

As expected, no differences were found between groups during the practice session. However, posttest performance was significant for Representation, $F(2,60) = 3.20$, $p < .04$, and Availability, $F(1,60) = 21.82$, $p < .001$. Scheffe' tests on the Representation factor ordered the groups, prose = flowchart > faded flowchart.

The repeated measures anova on original groups yielded significance for the Availability main effect, $F(1,54) = 12.36$, $p < .001$, and the Availability X Test interaction, $F(1.54) = 7.97$, $p < .006$. Across groups, the availability of the tax law was significantly more critical on the delayed test than on the immediate. When subjects were regrouped according to Availability for the delayed test, the Availability factor was again highly significant, $F(1,54) = 64.77$, $p < .06$. and the Representation X Availability interaction was significant, $F(2,54) = 3.37$, $p < .04$. Individual comparisons on the interaction showed that all three groups performed less well without the law. However, while there was no difference between groups who had recourse to the law, the faded flowchart group performed significantly worse than the other two groups when the law was withheld.

These results suggest that the type of task employed in the present experiment can be effectively taught through algorithms. Subjects were able to solve complex tax problems involving seven discriminators and six operators at a 75% level of accuracy after only 12-15 minutes of instruction. The algorithms required the learner to make a maximum of three unambiguous binary decisions which always lead to the same result. Although statistically significant, a decrement of only 15% occurred when two of the three algorithms were withheld following instruction, even after a one week interval. On the other hand, when subjects were allowed to retain the tax law during the testing session, no significant loss in accuracy was observed over time. In addition, analyses of the time data demonstrated a 100% increase in efficiency over the same period (Gerlach & Schmid, 1977).

Although one study cannot demonstrate conclusively which algorithmic form of instruction is most effective, these data do provide some tentative answers to this crucial problem. The prose and flowchart treatments performed consistently better than the faded flowchart group, despite the lack of differences in performance during practice. The similarity between these two conditions was that they provided the subject with an easy, systematic means of solving relatively complex tasks without requiring piecemeal mastery of the procedure. Subjects utilized the algorithm as a problem solving device, and simultaneously learned the procedure, resulting in high effectiveness. The faded flowchart was similar to traditional classroom instruction in that subjects were taught one section of the algorithm, asked to master it, and then go on to the next section. Such a system denies the step-by-step, logical characteristics of algorithmic instruction, and utimately proves less effective and less efficient. The Availability data also suggested that when algorithmic information is faded out during instruction, subjects perform significantly less well over time than when the algorithm is dealt with in a more unitary, organized fashion.

Educational Importance of the Study. The study indicates that, when dealing with instruction involving application of rules to solve computational problems, algorithmized instruction is effective. Rather than having the learner attempt to internalize the algorithm, the instructional designer should make the algorithm available during instruction.

During this study we also studied the efficiency of the algorithmized instruction. Our goal was to assess the efficiency, in terms of both instructional and posttest performance items, of an instructional algorithm in three representational forms (prose; flowchart; gradual withdrawal of flowchart), in two availability modes (with recourse to algorithm during practice and posttest; without recourse). The same design, materials, and procedures were used.

Data Source. Data consisted of time on posttest for 77 undergraduate volunteers from Arizona State University. Eleven subjects were dropped from the experiment for failure to follow procedures, leaving 11 per factorial cell. Subjects were run in groups during normal classroom sessions. Materials were prepared beforehand and shuffled, so that assignment to treatments were completely randomized.

Results and Conclusions. Time data was generated by computing the mean number of seconds spent per problem during instruction and during testing. Omitted problems were not included in the estimates.

Analyses of variance (3 Representation X 2 Availability) were conducted on the practice and posttest sections. A repeated measures anova using the delayed test data was then performed. The repeated measures anova was conducted twice: first, using groups as represented by the with/without distinction on the immediate posttest, and then using groups separated according to the Availability factor on the delayed test.

Time values obtained during the practice section were significantly different for Representation, $F(2,60) = 8.44$, $p < .001$, with Scheffe' tests ordering the means, flowchart < prose = faded flowchart. However, immediate posttest times did not differ significantly, either for the three algorithm groups or for the availability of the tax law. Subjects who had recourse to the law did not solve the problems more quickly than did subjects without recourse to the law.

The repeated measures anova on the original groups yielded significance for the Representation factor, $F\ (2,54) = 3.17$, $p < .04$, and for the Test main effect, $F(1,54) = 15.28$, $p < .001$. Scheffe' tests indicated that prose and flowchart groups spent significantly less time than the faded flowchart group. Less time was spent on the delayed test than the immediate posttest. When subjects were regrouped according to Availability for the delayed test, Representation was significant, $F(2.54) = 3.83$, $p < .02$, the Test factor was significant, $F(1.54) = 11.98$, $p < .001$, and the Representation X Availability Interaction reached significance, $F(2,54) = 6.16$, $p < .004$. Scheffe' tests again showed that the prose and flowchart groups spent less time than the faded flowchart groups. Subjects also spent significantly less time on the delayed test. Individual comparisons on the interaction demonstrated that only the faded flowchart groups spent significantly more time when not supplied with the tax law. No other differences were found.

These results suggest that the type of task employed in the present experiment can be efficiently taught through algorithms. Subjects were able to solve complex tax problems involving seven discriminators and six operators at a 75% level of accuracy after only 12-15 minutes of instruction. The algorithms required the learner to make a maximum of three unambiguous binary decisions which always lead to the correct result. Surprisingly, subjects spent significantly less time solving problems after a one week delay, with a decrement of only 15% when two of the three algorithms were withheld.

Although one study cannot demonstrate conclusively which algorithmic form of instruction is most efficient these data do provide some tentative answers to this crucial problem. The prose and flowchart treatments completed the problems consistently faster than the faded flowchart group, despite the fact that the faded flowchart group enjoyed significantly more practice time. However, overall efficiency was greater for the flowchart group than the faded flowchart condition. The efficiency of the faded flowchart continued to diminish over the one week delay. The similarity between the prose and flowchart conditions was that they provided the subject with an easy systematic means of solving relatively complex tasks without requiring piecemeal mastery of the procedure. Subjects utilized the algorithm as a problem solving device, and simultaneously learned the procedure, resulting in high efficiency. The faded flowchart was similar to traditional classroom instruction in that subjects were taught one section of the algorithm, asked to master it and then go on to the next section. Such a system denies the step-by-step, logical characteristics of algorithmic instruction, and ultimately proves less effective and less efficient. The Availability data also suggested that when algorithmic information is faded out during instruction, subjects performed significantly less well over time than when the algorithm is dealt with in a more unitary, organized fashion.

Educational Importance of the Study. The study demonstrates that, when dealing with instruction involving rule application to the solving of computational problems, algorithmized instruction yields a significant time savings. Of particular importance for instructional developers is the indication that, rather than trying to teach the learner to internalize an algorithm, the Learner should have constant recourse to it; both speed and accuracy are increased under this condition.

As a result of the above research, we next turned our attention to a study of structure in algorithmized instruction. An algorithmic procedure was examined under several instructional approaches pertinent to military training. A version of a logical, familiar, and computational algorithm was generated, further divided into verbal and symbolic (mathematical) forms, presented to sixty students, and tested for effectiveness and efficiency immediately following instruction and one week later. Multivariate analyses were employed. Learners achieved high scores following very brief instruction, with verbal and logical groups outperforming their comparison treatments. Several principles for use of algorithms in instructional design are suggested. Retention losses after the time delay were found to closely match the typical needs of a training environment.

These efforts led us to question the role of feedback in the type of instruction under study. We then designed an experiment to determine the effects of three levels of feedback (information, KCR, or none) on the mastery of an algorithmic procedure. Subjects were 63 eighth grade students. Treatments were delivered via self-instructional programs. A 16-item criterion referenced posttest was used. Results showed no major differences for the three levels of feedback on acquisition. While the results apparently confirm much previous research, lack of significance for feedback (of either type) compared with no feedback is unusual. Further research is needed to determine the most appropriate type of feedback for algorithmized instruction.

b. Performance Objectives.

For several years, we have been working on the development of a set of rules for writing performance objectives which can be demonstrated to function in a specified manner. During the past year, we re-analyzed nearly all of our previously gathered data, synthesized six separate studies which we had conducted, and, as a result, generated several new interpretations. Since reports of the research exist in the several technical reports submitted, we will limit our final report to those ideas and concepts which are new. Furthermore, all studies have been compiled in a single Technical Report, "Performance Objectives", TR#81230, December 1978.

We believe that our research has demonstrated that while contemporary instructional systems assume the use of some form of performance objectives, empirical support to justify their use is lacking. The vast number of studies on the topic demonstrate a contradictory three-way split among achievement increment, decrement, and no effect. However, the intuitive desirability of objectives and scattered positive affective data have provided objectives the support they need for continued use.

The research and rational exercise contained in this report offer a partial explanation of why such confusing results have plagued efforts to demonstrate the effect of objectives. It was first contended that neither the function nor the form of performance objectives have been operationally defined. Of the studies reviewed, ambiguous definitions eliminated the possibility of replication with even minor generalization. The dependent measures also varied widely, ranging from teacher behavior to student affect and achievement, precluding the possibility of any practical conclusions (save those regarding research methods). Thus, research has yet to provide an empirical demonstration of objectives' function.

As to the form of objectives, all agree that it is necessary to state explicit behaviors, conditions, and criteria (standards). Argument begins in determining what is considered explicit. Researchers addressing this issue have attempted to compile "recommended lists" of verbs to be used in objectives, and it is this kind of data which provided a springboard for studies which were designed to examine the form of objectives.

Our first efforts supplied us with normative data regarding the "observability" of 99 frequently recommended verbs, a replication of Deno and Jenkins (1969). The results confirmed their findings, but carried with them, as had previous studies, the implication that the verb is the only part of the objective worth considering.

Next we began an examination of the effect of <u>other</u> parts of the objective, i.e., conditions and/or criteria, on the already established "observability" of the verbs used previously. Both components were found to affect the nature of the objective as well as the perceived precision of the verb, often in a highly significant manner.

Then we further developed the role of the three components of performance objective. The data suggested that the observed precision of an objective results from an interaction of all three parts, and that anlayses of the components in isolation are meaningless.

The search for unaccounted-for variance in the overall observability of objectives led to an examination of yet another segment, the direct object. The objectives employed in all our previous studies used an abstract direct object (x or y) in order to minimize differential effects. In the next study, direct objects varying in judged degree of specificity were incorporated into the objectives and tested for their specific and general effects. The direct objects were found to contribute important additional information to the reader in terms of both observability and precision, following the linear relation anticipated by the gradations in specificity. Interestingly, comparisons of objectives with either abstract or concrete direct objects produced no overall significant difference, suggesting that readers had naturally substituted concrete modifiers for the abstract direct objects in order to "complete" the objective. An acceptable performance objective, therefore, should contain not only a verb denoting an observable behavior, but also appropriate conditions, criteria, and clearly stated or easily inferred direct objects.

As reported last year, a study was designed to determine whether or not the previous results could be extended to a specific target population -- that of military personnel (both trainees and trainers). Quite simply the similarity of the military personnel responses to those of the earlier studies (college students) on the verb rating tasks strongly supported the generalization of conclusions.

Further information on similarity of interactive effects will be required to provide a blanket acceptance of transferability.

In the spring of 1978, we designed a final study. In one sense, it turned out to be another ill-fated attempt to surface an observable function of objectives. As Illich (1973) so tactfully warned, "...alchemists failed no matter how often they tried, but each time their 'science' yielded new reasons for their failure, and they tried again." Indeed, through the pale of nonsignificance, a hopeful glimmer did occur. The observers in this study were asked to pass judgment on the "clarity" with which the teacher disseminated the information. While one supposes that teachers employing objectives would behave with corresponding precision, such was not the case. Nevertheless, it was discovered that the trained observers could effectively identify those groups which would more likely achieve. The presence of performance objectives appears less important than the behavior assumed to be associated with the use of objectives. While this finding is by no means surprising, it may suggest that researchers should abandon the methods currently in vogue in studies designed to ascertain the function of objectives in instruction. It is unlikely that any study will profoundly alter the basic teaching behaviors of the subjects. Even if a researcher did take great pains to provide a comprehensive training, there would be no guarantee that the comparison group would not have received similar "objective-based training" by conventional means.

To continue elaborating on the above would force us to go beyond the data. Thus, clearly demarcated, the following brief analysis is a rational defense for the use of performance objectives.

The complaints which reasonably well-motivated students most often make about courses fall into three main categories: (a) they do not understand what is going on, (b) they perceive the course as being too much work, and/or (c) there are components in the instructional environment which are either frustrating or unjust. Objectives provide a direct and positive means for reducing or eliminating all three. The use of objective-based training, or the systems approach, automatically excludes the first complaint because the student either lacks the necessary prerequisite knowledge and is denied admission until this deficiency is removed, or the "objectives" themselves are ambiguous or incomplete, a circumstance which the instructor must rectify. The systems approach also addresses the second class of complaints: all good instruction is learner-paced, or the instructional system must be adaptive to the learner. In addition, objective-based instruction tends to be simpler (though we have no empirical evidence for this assertion), a phenomonen which may be a byproduct of clearer thinking. Furthermore, this "simple" quality would obscure sought-after differences in empirical research. Finally, performance objectives help eliminate frustration and unfairness because of

(a) reduced test threat, (b) apparent concern for the student by the instructor (affective), (c) determination of a starting point, content, and a relizable end, and (d) equal requirements (explicitly stated) for all.

The question remains, "Do objectives facilitate learning?" The rational response would be, "...more research is not needed to establish a clear positive relation between objectives and good instruction." However, both rational thought and the empirical evidence presented in this report suggest that work is needed in the area of teacher or instructor training to promote the behavior induced by the effective use of performance objectives.

c. Self-evaluation and Performance Measurement.

In an effort to maximize cost effectiveness of training programs, all the military services currently use "objectives based" instructional programs for training. Trainees are told what performance is expected of them and given the criteria by which the performance will be judged. It is assumed they will use the criteria to evaluate the quality of their own performances and to alter performances which are not as they should be.

When job performance, rather than test scores, is at issue, the ability to self-evaluate takes on increased importance. If individuals are unable to evaluate their own job performance, if they are unable to discriminate between a correct and an incorrect performance, the results of training will be less than optimal. The question to be considered is whether trainees are able to use objectives and criteria to evaluate their own performances, and if so, whether the ability to self-evaluate improves with practice.

Experimental situations were created to investigate the effects of two independent variables: (a) learners' self-evaluation of their product using specified criteria and (b) learners' self-evaluation with instructor feedback on the accuracy of the evaluations; on two dependent variables: (a) learners' ability to use stated criteria for judging their own products, and (b) learners' performance on course objectives.

Literature on systematically designed instruction and the variety of instructional strategies that can be employed to improve learning is abundant. For the purposes of this study, literature that contained descriptions of (a) systematically designed instructional procedures related to objectives, standards, and practice and feedback and (b) self-evaluation as an instructional strategy was reviewed.

Instructional Strategies. We first explored literature on systematic instructional design procedures and the effects of using systematically designed instruction. Particular areas studied were performance objectives, performance standards, and practice and feedback.

A number of learning theorists state that learning will improve as a result of pre-specified objectives (Gagné & Briggs, 1974; Gerlach & Ely, in press; Kibler, Barker & Miles, 1970; Kibler & Basset, 1977; Mager, 1961). More research is needed on the development of strategies that will increase the ability of students themselves to use objectives in learning situations. One question about objectives which remains unanswered is whether students can effectively use only statements of objectives for maximum instructional effectiveness and, if not, what other instructional techniques preceding instruction can be employed along with objectives to facilitate learning.

The specification of performance standards is widely accepted as an integral part of an instructional objective. When performance evaluation is judgmental, when there is no clearcut right or wrong performance, the specification of performance criteria becomes appropriate. When a simple right or wrong response is not possible, Dick and Carey (1978) suggest that instructional objectives should include a checklist of the types of behaviors which will be expected when the performance of students is judged. This should give students a clear understanding of the nature of the required performance. However, the degree to which learners are able to use criteria in shaping their behaviors has not yet been established (Carey, 1976).

The use of practice and feedback are universally accepted as important aspects of instruction. Both are regularly included in systematically designed instructional materials. Research on practice and feedback were reviewed in this study relative to self-evaluation or student evaluation of their own work. Kulhavy, Yekovich and Dyer (1976) have shown that, in programmed instruction, certain conditions maximize the effects of instructor feedback on learner performance. They found that feedback in programmed instruction is most effective when a learner believes a test response is correct when, in fact, it is incorrect. They say that, in this case, learner response confidence is incorrectly high. When a learner believes a response is incorrect, and it is, Kulhavy et al. state that feedback is not effective because learners do not understand either the subject matter being tested, the test question, or both.

In programmed instruction, feedback indicates to the learner whether a particular response to a program frame is correct or incorrect, and the effects of feedback are measured by a subsequent test covering the same material. In some instructional situations, the feedback given on one performance is expected to help the learner with future performances which are similar. When complex, multi-dimensional performances are given, learners' confidence in this performance can be measured by having them evaluate each aspect of their own performance. Self-evaluation becomes an indicator of learner response confidence. If instructor feedback is then given on the quality of learner self-evaluations, in addition to the performance of tasks to be learned, a double feedback condition exists. The effects of feedback on student performance should be noticed under these conditions regardless of learner confidence, especially when the feedback given on performance and on self-evaluations is instructional and causes students to attend more to objectives and standards for acceptable performance.

Self-evaluation. One method for determining the degree to which learners are able to use objectives and criteria statements to shape their performances is to ask learners to use them to evaluate their own work. Those who are able to evaluate their work using pre-specified criteria probably possess a clear understanding of the objective and criteria concerned. On the other hand, learners who are unable to use performance criteria to judge the quality of their work after its completion are probably unable to use the same criteria during its production. Learners who are unable to use criteria during performances would not be expected to perform as well as those who have the ability to apply statements of criteria when developing instructional products.

Clark (1938) found that students were able to evaluate their own performance on college level algebra, quantitative analysis, and chemistry problems. The correlation between student scores and instructor scores was .80. Bennent (1958) supports the argument that students are able to evaluate their own performances.

Both Clark and Bennent report that students gave themselves the same grade as their instructors in a high percentage of cases. Estimating a single grade, or single number on a 5-point scale seems to be an easier task than determining how closely several parts of a complex performance match the related performance criteria. When one performs in the field, grades are not a consideration, job performance is, and evaluations are based on how closely the actual performance approximates the expected performance.

The question of whether self-evaluation actually promotes learning remains unanswered. McEowen (1957) found that learning was unaffected by students' evaluation of their materials. In this study, however, performance had no effect on course grades, so students had little at stake. Had the performance affected the grades, student interest in learning might have been greater and the results might well have been different.

In a study involving sculpture, graphics, painting, and drawing skills, Fried (1965) found that although sculpture improved as a result of self-evaluation, the other skills studied did not. Noting the inconclusiveness of his own finding, Fried concluded that the value of self-evaluation was still undertemined, and recommended continued research in the area.

In a study conducted in two Air Force technical schools, Duel (1958) found that achievement was improved when students were given formal and periodic opportunities to evaluate their own work. A study in which self-evaluation had a positive effect on students' mechanical drawing skills was conducted by Irwin (1973). He found that students who evaluated their own materials throughout the school year learned more as measured on a standardized test which assessed mechanical drawing skills. In this study, however, results were confounded because the self-evaluating group had a significantly higher mean IQ than the control group.

Research into the effects of self-evaluation has generally been concerned with either student ability to self-evaluate, or the effect of self-evaluation on student performance. An indication from the literature is that, in certain cases, students are able to evaluate their own materials, although the evidence is far from conclusive. Evidence has been cited which both supports and contradicts the assertion that self-evaluation improves student performance. Little has been written on the effects of (a) student practice on evaluating their own work and of (b) instructor feedback on student self-evaluation. The combination of systematic instructional procedures with self-evaluation as an additional feedback strategy may prove to make instruction more effective.

Hypotheses. The literature concerning learner self-evaluation, though limited, does indicate that students may possess at least a limited ability to evaluate their own work and that periodic self-evaluation may improve learner performance.

The specific hypotheses tested in this research are the following:

1. The correlation between learner and instructor evaluation scores on a final course project will be higher when instructor feedback is given on previous learner self-evaluations throughout the term than when no instructor feedback on self-evaluation is given, or when students do not participate in self-evaluation throughout the term.
2. Learner performance on a final class project will be better when students evaluate their own materials throughout the term and receive instructor feedback on their self-evaluations than when students do not receive feedback on their self-evaluations, or do not participate in self-evaluation throughout the term.
3. For a subgroup of students who incorrectly assume that their initial products are well produced learner performance on a final product will be better when feedback is given on the quality of the students' evaluation of their own work than when instructor feedback is given only on the quality of the product.
4. For a subgroup of students who correctly assume their product is not well produced learner performance on a final product will be better when feedback is given on the quality of the initial product and on student evaluations of the project than when feedback is given only on the product.

Sample. A total of 56 students enrolled in Audiovisual Materials and Procedures in Education at Arizona State University were used as subjects. Students were blocked by grade point average (high, medium, low) and randomly assigned to either the first treatment, second treatment, or control group.

Treatment. Control group students used performance objectives, criteria statements, and instructor feedback to produce a mediated unit of instruction. There were three major checkpoints during the semester at which students formally submitted their materials for instructor evaluation and feedback. The self-evaluation group followed procedures previously described for the control group. In addition, prior to submitting their materials for instructor evaluation, they used the instructor evaluation form to assess the quality of their own work. Instructor feedback on the quality of the materials was the same for both the control and the self-evaluation groups. The third or feedback group produced their materials in the same manner as the other two groups. Like the self-evaluation group, prior to submitting their materials, they evaluated their own work and submitted the evaluation forms to their instructors. Instructor feedback for this group consisted of both feedback on the quality of the products and on the quality of the student evaluations.

To enable comparisons in self-evaluations among the three groups, the control group used the evaluation forms to assess their final products at the end of the semester.

Performance Standards. Each aspect of student products was graded by course instructors on a scale of 1 to 10, with 10 being the highest score possible. Grading criteria were established by course instructors prior to the beginning of the study. Instructor scores were used as the standard against which all comparisons were made.

Design. In this study a posttest-only control group design was used to measure the effects of student self-evaluation of instructional products, and instructor feedback on the quality of self-evaluation on student performance on the final class product. An analysis of variance at a significance level of .05 was used to measure the effects of these two independent variables. A Pearson product-moment correlation at a significance level of .05 was used to measure the relationship between student evaluations and instructor evaluations of all instructional products.

Results. Self-evaluation and feedback on the accuracy of self-evaluation were studied to ascertain their effects on (a) student's evaluation of their own performance and (b) on student achievement.

It was hypothesized in this study that the correlation between student evaluation scores and instructor evaluation scores on the final course project would be higher when instructor feedback on student evaluations of other class projects was given than when no instructor feedback was given, or when no student evaluation took place.

For the control group, the student evaluation mean was 57.9 (maximum score = 60) and the instructor mean was 56.82, yielding a difference of 1.08 points with the correlation between them being .68 ($p \geq .05$). For the self-evaluation group, the student mean was

58.5, with the instructors giving a mean score of 53.67, differing by 4.83 points. The correlation between these two scores was -.07, which was not significant. Students in the feedback group predicted a mean score of 55.70 while the instructors' mean equaled 54.75, yielding a difference of .95 points. The correlation between these scores was .81 ($p \geq .05$).

There was very little change in student ability to evaluate their own work in the self-evaluation group. At the first checkpoint, the correlation between student and instructor scores was -.048 and at the third checkpoint, -.076. Neither of these correlations was significant. There was little association at the outset of the study and only slightly more at the conclusion.

On the other hand, the correlation between student and instructor scores for the feedback group was .49 ($p < .05$) at the first evaluation checkpoint and .81 ($p < .05$) at the third. The relationship between the two sets of scores for this group increased approximately 65 per cent.

It was predicted that students' performance on the final class project would be better when students evaluated previous class products and received instructor feedback on the self-evaluations than when they completed self-evaluations alone or did not complete self-evaluations at all. A one-way analysis of variance revealed no significant difference between group means.

A one-way analysis of variance indicated that when student confidence was incorrectly high, there was a significant difference between the self-evaluation and feedback groups $F = 9.03$ $(1,21)$ $p \leq .01$. When student confidence in materials was correctly low, the feedback group mean was approximately six points lower than the mean score of the self-evaluation group.

Discussion. The purpose of this study was to investigate the effects of (a) student self-evaluation and (b) instructor feedback on the self-evaluations on the ability of students to evaluate their own materials and on their performance. Although criteria statements are often included in performance objectives, there has been little evidence which indicates that students can always use criteria to shape their performance (Carey, 1976). The instruction used in this study included objectives with criteria statements and relevant practice and feedback. It conformed to the prescription for well developed instruction as specified by Gagné and Briggs (1974), Dick and Carey (1978), and Gerlach and Ely (in press).

The relationship between student score predictions and instructor scores on the final class product was reasonably high for the control group, $r = .68$ ($p \leq .05$), supporting the work of Clark (1938) and Bennent (1969). However, the combination of self-evaluation and instructor feedback on self-evaluation raised this figure to .81, indicating that the skill of self-evaluation can be developed. These findings support the hypotheses stated in this paper. Criteria specific feedback on self-evaluation seems to be an effective way to develop student skills in estimating the quality of their performances.

The correlation between student and instructor evaluation scores for the control group was much higher at the third evaluation checkpoint than the correlation for either the self-evaluation or feedback group at the first checkpoint. For the control group the final product was their first opportunity to evaluate their own work. One possible explanation for this phenomenon is that as students go through a course, they acquire an understanding of instructors' expectations and how well their performances meet these expectations. Students then adjust their performances accordingly.

It was anticipated that self-evaluation itself without any instructor feedback on the accuracy of the evaluation would raise the correlation between student and instructor scores. The analysis revealed that self-evaluation itself had no beneficial effects, and in fact, seemed to have a negative effect. Self-evaluation seemed to confuse students and hamper their ability to use prespecified criteria to assess the quality of their materials. We have no explanation for the occurrence of this effect in either the pilot study or this study. There is no literature on which to base an explanation of this kind of result.

The effects of self-evaluation and feedback on student performance is difficult to gauge in this study. Students in all three groups achieved a grade of A on the final project scoring an average of 55 out of a possible 60 points. There was little variability in student performance.

When tied to student confidence in their materials, there was a significant difference between group means. Feedback on self-evaluation of earlier material seems to be highly effective in promoting improved student scores on later performances when student confidence in materials was incorrectly high. When students received instructor feedback on the quality of their materials and on their own evaluations, they had the opportunity to learn not only what instructors thought of their materials, but how their self-evaluations compared to instructor evaluations on an objective by objective basis. This technique pushed the mean scores of students in the unwarranted high confidence subgroup 3.5 points above the feedback group mean from 54.75 to 58.25. This tends to support the conclusions of Kulhavy, Yekovich and Dyer (1976) that feedback is most effective when a high degree of student confidence in materials is unwarranted. In the self-evaluation group, the difference between the mean scores of students who had incorrectly assumed their materials were good and the entire self-evaluation group was minimal, lending more support to the conclusion that the addition of feedback on the self-evaluations is beneficial when learner confidence in their work in incorrectly high.

When student confidence in materials was correctly low, feedback on self-evaluation seemed to have a negative effect on performance, lowering the subgroup means from 54.75 to 49.50. These results contradict the hypothesis stated earlier that feedback on self-evaluations is effective for all students, but support the findings of Kulhavy, et al. who concluded that when student confidence was correctly low, they were confused, and feedback was not effective. The results of this study indicate that when students were unsure of their work, additional feedback only caused more confusion.

Duel (1968) and Irwin (1973) both concluded that self-evaluation did improve student performance. Due to the ceiling effect noticed in this study, the current results can neither support their conclusions, nor refute the findings of McEowen, who found no improvement, and Fried (1965), who found improvement in some situations. It is clear that the effects of self-evaluation on student performance would possibly be more informative under conditions in which student grades are normally distributed.

Although the research seems to support the results of the Kulhavy et al., (1976) research concerning the effects of feedback, continued research in this area is also indicated in order to clarify the effects of feedback in learning situations in which student performances are complex and student confidence is difficult to gauge.

There were many questions raised during this study that will need to be investigated in future studies. Two were of particular interest. The first question relates to the unexplained negative effects of self-evaluation only on student evaluations of their own work. The researchers expected this activity to increase the ability of students to assess their own work. When instructors appeared to ignore student evaluations, students were not helped.

A second question of interest is where to begin instruction for students who are unable to use prespecified criteria to improve the quality of their work. If students cannot perform specified tasks, cannot use objectives and criteria statements to evaluate their own work, and cannot use instructor feedback to improve their products, then the problem may be that these students do not possess prerequisite skills for the instruction. A feedback strategy that includes information on current instruction and on prerequisite skills instruction should be investigated in an attempt to identify the best strategy for this particular subgroup of students.

If trainees can learn to use effectively prespecified objectives and criteria effectively to evaluate and shape their performances, the case for the use of instructional objectives is made even stronger. Then, not only will trainees know precisely what is expected, but they will be able to tell if and when they are able to perform the requested tasks.

If trainee self-evaluation can be developed to an acceptable level, then time spent on instructor feedback can be greatly reduced, freeing trainers for other instructional tasks during the training process.

d. Algorithmic Instruction and Computer Models.

Studies were conducted (1) to examine the relationship between computer modelling and human algorithmic behavior, (2) to develop a set of predicted learning outcomes based on this relationship, and (3) to test the validity of the logic with empirical data.

Theoretical framework. The use of computer modelling as an explanatory vehicle for psychological processes is a widely used and productive practice (Anderson and Bower, 1973; Feigenbaum, 1963; Newell and Simon, 1961; Quillian, 1968). The development of algorithms in instructional theory has also benefited both directly and indirectly from computer science. Computer-type programs, which are by definition "pure" algorithms, have been directly integrated into the classroom as teaching devices (Landa, 1966; Scandura, Durnin, Ehrenpreis, and Luger, 1971; Schmid and Gerlach, 1977; Schmid, Portnoy, and Burns, 1976). The indirect use of computer models via the science of cybernetics has assisted theoreticians in their formulations of algorithmic learning theories (Bung, 1968; Frank, 1964; Landa, 1966; Lansky, 1966). A logical extension of this scientific evolution is the use of computer algorithms and computer language and terminology in describing human algorithmic behavior. Following is a very brief description of the model's terminology and its structural implications for instruction.

Structure of the model. Essentially four basic computer terms are employed and described within the context of algorithms. The first two terms, width and depth, refer to the number of branches in a complete algorithm (width) and the length of each branch (depth).

Instructional and learning effectiveness and efficiency can be facilitated based on the correct use of these characteristics. For example, if Miller's (1956) principle of 7 ± 2 is applied to the "complexity limit" of an algorithm to be taught in the classroom, both the instructional designer and the learner are likely to benefit. In addition, it has been pointed out by Landa and others that there are usually many possible variations for an algorithm designed to teach a given content or process. An algorithm providing the minimum depth and width given the learners' entry skills could prove to be the "best" algorithm instructionally.

The other two terms are serial and parallel processing. When computer hardware contains more than one processor of the same operation, or processors of several different operations, maximum efficiency can be yielded when the program calls for the processing of several operations in parallel (simultaneously), given that the problem solution allows it. Both learned and unlearned human psychomotor behaviors may apply to this characteristic. Human cognitive processes, on the other hand, may be better likened to computer time-sharing where, at a given level of analysis, the computer deals with operations one at a time, and often retains subsolutions for use in subsequent operations. In such cases, a sort of control structure is necessary in a computer to coordinate the separate activities, and it can be assumed that human cognition also requires a similar coordination of processing. (While the control structure also carries behavioral ramifications for algorithmic instruction, it is beyond the scope of a single study to pursue them.) The instructional question which arises from the serial/parallel distinction is, "What is the optimal strategy for presenting a branching algorithm?" Most classroom or training tasks involve several serial progression of steps which always lead the learner to the correct solution. The present study examines the data of two studies designed to begin exploring the problems outlined above.

Method and results. The algorithm, used to calculate tax on stock transactions (Schmid and Gerlach, 1977), consisted of five discriminators and five different operators. In terms of the model, the "width" of this algorithm was five: five operators or discriminators on the same level. The "depth" of this algorithm was four, because the longest branch consists of four serial transitions.

Study One provided a test concerning parallel processing: are algorithms better taught in serial subcomponents, in parallel clusters, or in a combination of the two? Subjects in Treatment One received a prose version of the algorithm and were allowed to develop their own solution strategy. Treatment Two subjects were given an intact flowchart and directed through the practice problems in an essentially serial fashion (solving the problems by using one complete branch at a time). Treatment Three subjects received the Treatment Two flowchart, but were forced to memorize its content in a top-down fashion, with information from upper levels deleted first. The flowchart was therefore treated as a single solution unit. Seventy-seven undergraduates participated, solving six practice problems and six test problems at their own pace. Achievement scores and times served as dependent measures.

The results of this study were quite dramatic. Students in Treatments One and Two performed equally well at a high level of mastery following only 12-15 minutes of instruction (approximately 85%). However, the group that was asked to learn the algorithm as a single unit performed significantly less well on both immediate and one-week delayed tests (approximately 60%). It was assumed that due to the similar achievement of the prose and flowchart groups, those subjects were working with the same algorithm and adopted the same strategy. To accept this assumption, however, one would expect Treatment Two to work more efficiently because the instruction directed them immediately to the superior strategy. The time data confirmed this hypothesis. The flowchart group mastered the algorithm in significantly less time than either of the other groups. The prose group, once it adopted and learned the proper strategy, worked as efficiently as the flowchart group on the posttests, both of which worked faster than the third treatment. Finally, one might expect the prose group to have learned and retained the strategy better because their treatment forced them to formulate tnd utilize the serial strategy without formal prompting. The significant Representation (the three instructional treatments) x Availability (the algorithm was withheld or made available during the posttests) interaction on time data confirmed this prediction.

Serial processing entails a string of transitions, each serving as a link in a single chain, each drawing from and contributing to the appropriate adjacent transitions. Serial processes are therefore always intrinsically logical and orderly. However, the logical order of things is lost to the computer. Its only equivalent to entry skills is the amount of time a processor requires to complete a given operation. Practically speaking, the computer either knows or it doesn't know, which means we, as users, either receive the answer very quickly with absolute accuracy, or not at all. Human thought usually falls somewhere between those extremes. To examine these shades of difference, we introduced human logic into an otherwise completely controlled algorithmic procedure in the second study (Schmid and Gerlach, 1978). For example, the original algorithm first asked, "Is the selling price greater than the market value?" A second algorithm solving exactly the same tax problem began with the question, "Did you make or lose money...?" The assumption was that the second question is more logical, thus easier to remember. This "logical" algorithm was written to possess the identical structural characteristics of the original: width = 5; depth = 4; entirely serial, employing seven discriminators and five operators. The logical algorithm was a graphic mirror image, changing only the "human logic" by means of content alterations. To further test the intrinsic value of the logical reformation, a third treatment, employing both the original and logical treatments, was designed using only letter symbols (e.g., "Is s > M?" or "Is the selling price greater than the market value?" and so on). The subjects were 60 undergraduates, who followed the same basic procedure as that used in Study One.

The Sequence (orginal vs. logical) main effect reached marginal significance for combined immediate and delayed posttest scores, and significance at the .02 level on the immediate test alone, both in favor of the logical algorithm. There were no confounding interactions with the Form (verbal vs. symbolic) factor. The time data provided even stronger differences, with both combined and separate posttest times in favor of the logical group.

Educational Implications. This demonstration provides tentative support for two points. (1) If learning is in fact facilitated by a serial approach (and the vast quantity of mnemonic research suggests that it should), then teachers would be advised to construct algorithms of a restricted width when possible. A pilot study also included in Schmid and Gerlach (1977) demonstrated that an extremely lengthy serial array becomes unwieldy for the learner. (2) Furthermore, parallel transition points are not likely to be similar enough for effective chunking. The same principle can be implemented in developing serial (depth) strings.

These results verify Landa's basic instructional principle of breaking the algorithm down and teaching it in logical parts until mastered. They also demonstrate that computer language can be used to describe precisely algorithmic behavior which may be useful to instructional researchers.

e. A Note on the Use of the Pearson Product-Moment of Correlation

When the range of one or both variables in a simple Pearson product-moment correlation is artifically restricted, the absolute value of the sample correlation coefficient is reduced below that of the true (population) coefficient. This phenomenon is responsible for the well known difficulty of finding useful correlations between grades and other indices of intellectual performance in such preselected populations as college students -- especially graduate students (cf. Wallach, 1976).

Restriction of range is expressed relative to the population, with the most widely accepted index being the ratio of the sample standard deviation to the population standard deviation (Ghiselli, 1964; Guilford, 1954). Ghiselli has presented the mathematical proofs which establish the required correction for attenuation when the population variance is known or can be estimated accurately. In the limit, as the ratio of the sample standard deviation to population standard deviation approaches zero, the Pearson correlation coefficient approaches zero.

What is rarely made clear in classroom discussions is that correlation is not merely a function of the marginal distribution but of the conditional distributions as well. Perhaps for this reason, there appears to be a widespread misunderstanding of this phenomenon, leading to the belief (until recently shared by us) that any variable with a variance near zero would necessarily show a correlation coefficient near zero with any other variable, and that decreasing variance necessarily leads to decreasing absolute values of coefficients. For this reason, it seems useful to describe an anomaly recently encountered in our research.

The Anomaly. During the course of our research on flying training performance objectives, we collected ratings of the "observability" of training objectives using a standard 1 to 5 rating scale. In addition to ratings of complete statements of objectives, we also collected ratings of the three component parts of such statements -- the condition, verb, and criterion (e.g., "given a pair of 9-digit numbers/add the two numbers/without error"). For a surprising number of statement components, there was near-universal agreement among the subjects participating. For example, all subjects but one rated "with 10% accuracy" as being highly observable -- a 1 on the rating scale. On scanning the data, we were confident that the computed correlation coefficient would be near zero, because of the small variance; the discovery of a coefficient of +1.00 came as a distinct surprise.

In retrospect, it is clear why the coefficient must be 1.00: a straight line will fit the data perfectly, and increases or decreases in the variance resulting from the addition or elimination of pairs falling on either data point (1,1 or 2,3) will not change the data pattern. The correlation coefficients computed for varying numbers of 1,1 pairs and extra selected data points show that addition of extra subjects at data point 1,1 decreases the variance but increases the correlation coefficient. The actual coefficients are dependent on the arbitrary selection of the extra points are chosen.

Conclusion. When small numbers of subjects are used with discrete rating scales of limited range, it seems likely that data patterns of this type will often be found. If all subjects use the same category, that rating necessarily correlates zero with any other scale. If one or two subjects choose divergent responses on each of two scales, the correlation may be quite high, even perfect, regardless of the variances obtained. Thus, the belief that decreased variance in one or both variables is always associated with reduced correlation is clearly incorrect. As an instructional note, we recommend that classroom presentations of the restriction of range problem be structured to avoid leaving such an impression with the student. While the data analyst himself is not likely to be misled, the risk is that those who do not have access to the raw data may take such correlations at face value and assume great predictive power where none exists.

f. A Disappointment.

We had high hopes of performing an experiment to test an hypothesis, regarding performance measurement; this hypothesis, in essence an extension of the Shipley method (see our Technical Reports #40830, August 1974 and #60229, February 1976), would permit accurate estimation of the quality of a complex skill by means of the evaluation of a single (or a very few) critical variables.

During the first months of the grant period we made a number of unsuccessful attempts to obtain access to subjects and simulators at Williams Air Force Base. Unfortunately, the conflicts between the needs of HRL, the on-going training mission, and our own needs for experimental "purity," could not be reconciled. We then turned our attention to designing a study which would achieve a comparable goal through the use of on-hand data, specifically, data collected from 1974 to 1977 which was stored on computer tape.

Although this approach was filled with promise when we began, we encountered one obstacle after another. Finally, in early February 1979 (some five weeks after the project concluded) we were forced to abandon this effort with more disappointment than it would be appropriate to describe in a technical report.

Despite this failure, we feel that the Shipley method is potentially valuable and we do hope to return to a further study of the problem as soon as the present Arizona State University computer is replaced (perhaps, summer 1980).

C. Personnel

Principal investigator:
Vernon S. Gerlach

Senior investigators:
Richard F. Schmid
Lou Carey

Research associates:
Scott Herrington
Larry Israelite

Consultants:
Miroslav Valach
Robert Ney
Michael Carson

Clerical
Kristin Anderson